Original Article / Özgün Makale

MEDICINE PALLIATIVE CARE

DOI: 10.47582/jompac.769064

J Med Palliat Care 2020; 1(4): 83-92

Detection of CD40 polymorphisms and investigation of their relationship with possible diseases by bioinformatics methods

Biyoinformatik yöntemler yardımıyla, CD40 polimorfizmlerinin tespiti ve olası hastalıklarla ilişkinin araştırılması

Duygu Kırkık[1], Faruk Berat Akçeşme[2], Sevgi Kalkanlı Taş[3]

[1]University of Health Sciences, Hamidiye Medicine Faculty, Department of Medical Biology, İstanbul, Turkey
[2]University of Health Sciences, Hamidiye Medicine Faculty, Department of Biostatistics and Medical Informatics, İstanbul, Turkey
[3]University of Health Sciences, International Medicine Faculty, Department of Immunology, İstanbul, Turkey

Cite this article as/Bu makaleye atıf için: Kırkık D, Akçeşme FB, Kalkanlı Taş S. Detection of CD40 polymorphisms and investigation of their relationship with possible diseases by bioinformatics methods. J Med Palliat Care 2020; 1(4): 83-92.

ABSTRACT

Introduction: Cluster of differentiation 40 is a type I transmembrane protein present on B cells, macrophages, dendritic cells, and endothelial cells, which leads to cell activation, proliferation, adhesion or differentiation. Previous studies have been shown that cluster of differentiation 40 polymorphisms have an effect on some autoimmune diseases. The purpose of this study is to investigate all single nucleotide polymorphisms found on cluster of differentiation 40 gene and their relationship with possible diseases by bioinformatics methods.

Material and Method: In our study, while GeneMANIA was used to investigate the relationship between cluster of differentiation 40 gene with other genes, SIFT was employed to select sequences with similar functions as cluster of differentiation 40 gene. Exome cariant server was used for the detection of changes between populations and suspected single nucleotide polymorphisms. Mr single nucleotide polymorphisms Software was used to predict the effect of binding of 3'untranslated regions single nucleotide polymorphisms to miRNA. In order to analyze the untranslated regions on single nucleotide polymorphisms, UTRscan tool was used. UbPred was used for the estimation of the potential ubiquitination site on proteins, and Prosite was used to define the functional characterization of the protein domain.

Results: In our study, a total of 85 single nucleotide polymorphisms were found for cluster of differentiation 40 gene, and rs147677886, rs11569321, rs7273698, rs11086998, and rs139300926 were detected as suspected single nucleotide polymorphisms. Moreover, these single nucleotide polymorphisms may be associated with multiple sclerosis (MS), rheumatoid arthritis (RA), and Kawasaki disease.

Conclusion: Currently, there are no studies in the literature about single nucleotide polymorphisms of cluster of differentiation 40 gene that we detected by bioinformatics methods. In the future, we aim to evaluate this study experimentally in the laboratory and contribute to population-specific studies.

Keywords: Bioinformatics, CD40, SNP, mutation, *in silico*

ÖZ

Giriş: Cluster of differentiation 40 olgun B hücreleri, monositler, dendritik hücreler ve sinyal iletimi ile ilgili epitel hücrelerde bulunan ve hücre aktivasyonuna, çoğalmasına, yapışmasına veya farklılaşmasına yol açan tip I transmembran proteinidir ve cluster of differentiation 40; B hücreli kronik lenfositik lösemiler, lenfomalar ve bazı karsinomlarda eksprese edilir. Bu çalışmada amaç; cluster of differentiation 40'ın biyoinformatik yöntemler yardımıyla tek nükleotid değişimlerinin bulunması ve olası hastalıklarla ilişkilerinin araştırılmasıdır.

Gereç ve Yöntem: Çalışmamızda; cluster of differentiation 40'ın olası genlerle ilişkisini araştırmak için GeneMANIA, benzer fonksiyona sahip muhtemel ilişkili dizileri seçmek için SIFT, popülasyonlara özel değişikleri analiz etmek ve şüpheli tek nükleotid değişimlerinin tespiti için Exome Variant Server, 3'UTR tek nükleotid değişimlerinin miRNA bağlanması üzerindeki etkisini tahmin etmek için mr tek nükleotid değişimlerinin Software, tek nükleotid değişimlerde untranslated bölgelerin analizi için UTRscan, proteindeki potansiyel ubikuitinasyon bölgesinin tahmini için UbPred ve protein domainin fonksiyonel karakterizasyonunu tanımlamak için Prosite kullanıldı.

Bulgular: Çalışmamızda farklılaşma 40 geni kümesi için toplam 85 adet tek nükleotid polimorfizmi ve rs147677886,

rs11569321, rs7273698, rs11086998 ve rs139300926 şüpheli tek nükleotid polimorfizmleri olarak tespit edildi. Ayrıca, bu tek nükleotid polimorfizmleri multipl skleroz (MS), romatoid artrit (RA) ve Kawasaki hastalığı ile bağlantılı olabilir.

Sonuç: Biyoinformatik yöntemlerle tespit ettiğimiz 40 farklılaşma gen kümesinin tek nükleotid polimorfizmleri ile ilgili literatürde şu anda herhangi bir çalışma bulunmamaktadır. Gelecekte bu çalışmayı laboratuvarda deneysel olarak değerlendirmeyi ve popülasyona özel çalışmalara katkı sağlamayı hedefliyoruz.

Anahtar Kelimeler: Biyoinformatik, CD40, SNP, mutasyon, *in silico*

Corresponding Author / Sorumlu Yazar: Duygu Kırkık, Sağlık Bilimleri Üniversitesi, Hamidiye Tıp Fakültesi, Tıbbi Biyoloji Anabilim Dalı, İstanbul, Türkiye
E-mail / E-posta: dygkirkik@gmail.com
Received / Geliş: 24.07.2020 **Accepted / Kabul:** 24.08.2020

This work is licensed under a Creative Commons Attribution 4.0 International License.

INTRODUCTION

Cluster of differentiation 40 (CD40) belongs to tumor necrosis factor (TNF) superfamily, is an integral membrane protein receptor of antigen-presenting cells such as dendritic cells and follicular dendritic cells. Also, it is represented on the cell surface of hematopoietic progenitor cells, epithelial cells, carcinomas and B lymphocytes (1). CD40 gene is positioned on chromosome 20 (q12-q13.2) that codes a protein found on the surfaces of cells as a structural trimer complex (2). T cell-dependent immunoglobulin class switching, memory B cell development, and germinal center formation are one of the reponses among those that is madiated by CD40 (3). TRAF1, TRAF2, and TRAF6 are adaptor proteins that are attached to the TNFR receptor and probably TRAF5 proteins interact with TNFR resulting a signal transduction mediated by this interaction (4). This receptor-ligand interaction is thought to be required to activate microglial activation induced by amyloid-beta. Thereby, it might be considered as an earler indicator in Alzheimer's disease pathogenesis. According to the previous studies, there have been two alternative transcript variants of CD40, which are coding different isoforms (5).

IFN-γ is a primary activation signal for macrophages which is produced by CD4 Th1 cells. The secondary signal is created when CD40 molecule on T cells is interacted with its ligand, CD40L (CD154), on the macrophage cell surface. Thus, the activation level of macrophages is increased by the help of high amount of CD40 and TNF receptor expression on their cell surface. Destructive materials to microbes such as reactive oxygen species and nitric oxide are induced in macrophages as a result of increased level of activation, directing macrophages to digest ingested microbes (6). On the other hand, antigens can be presented by B cells to helper T cells. B cell activation is carried out through receptor-ligand interaction of CD40 receptor expressed on B cells and CD40L ligand expressed on activated T cells. Also, T cells generate IL-2 that directly affects B cells causing a net stimuli leading cell division, differentiation to plasma cells, and antibody isotype switching of B cells. At the end, B cells have the ability to make antibodies specific to target antigens (7). The constructive expression of CD40 is found on antigen-presenting cells including dendritic cells, macrophages, and B cells. Several normal body cells such as smooth muscle cells, endothelial cells, fibroblasts and epithelial cells express CD40 on their surface (8). A wide range of tumor cell expression of CD40 has been also reported including Hodgkin's lymphoma and non-Hodgkin's lymphoma, myeloma and nasopharynx, bladder cancer, cervical cancer, kidney, and ovarian cancer. B cell precursors in the bone marrow are another cell type in which CD40 is expressed and it has been demonstrated that CD40-CD154 interaction plays a role in the control of B cell hematopoiesis (9). New concepts have been revealed about the relationship between CD40 and the pathogenesis of different autoimmune diseases. Also, several loci that occupy an important position in the progression of various diseases were identified (10,11). CD40 locus was defined as a genetic risk factor for various diseases including rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) (12,13). Recent studies have been demonstrated that there are numerous polymorphisms identified in the CD40 coding gene, and CD40 gene polymorphisms trigger several autoimmune and inflammatory diseases including diabetes, fuchs uveitis syndrome, essential hypertension, Graves' disease, and cerebral infarction (14-18). Nowadays, genomic information and biological data can be interpreted and saved by bioinformatics methods. In this study, the investigation of single nucleotide polymorphisms (SNPs) in the CD40 gene and its association with possible diseases was aimed.

MATERIAL AND METHOD

In this study, in silico methods have been used and do not require any ethical committee approval. All procedures were performed adhered to the ethical rules and the Helsinki Declaration of Principles.

Analyzing of Functionally Similar Genes with GeneMANIA Software

GeneMANIA is a database allows analysing similar genes. It is a software for analyzing functionally similar genes with an input gene which is defined as (according to their) physical interactions, co-expression, co-localization, or genetic interactions. The software, where it can be reached from the web address "https://GeneMANIA.org", gives the user a result list by analyzing genomics and proteomics data. In this study, the CD40 gene was used as an input gene and its interaction with other genes was shown.

Using Exome Variant Server for Analyzing of Suspected Genes

Exome Variant Server database is supported by National Heart, Lung, and Blood Institute (NHLBI). The database helps to find new genes and mechanisms of heart, lung, and blood related diseases. The database can be reached from "http://evs.gs.washington.edu/EVS/" web address. Exome Variant Server is used to determine the number of variations according to population and it allows monitoring the number of EA (Population of Europe and Africa) allele and AA (Population of Africa and America) allele. In our study, suspected SNPs were identified for both CD40 and TNFSF4 genes and allele frequency were analyzed with this database. At the same time, AA and EA genotypes were calculated and evaluated.

Detection of Effects of SNPs on microRNA by Using mrSNP Software

mrSNP is a tool used to predict the effect of an SNP in 3'UTR on miRNA binding with high accuracy. There are several advantages of this tool compared to other available algorithms. This tool allows the user to enter any SNPs that have been characterized by any SNP searching program. Testing the performance of mrSNP on SNPs experimentally verified whether this is affecting miRNA binding, mrSNP defined 69% (11/16) of the SNPs that prevent binding correctly. The web address of the software is "http://mrsnp.osu.edu". In this study, SNPs were evaluated on the CD40 gene with this software.

UTR Scan

Untranslated regions on SNPs can be identified via the UTRscan program. The UTRscan is a pattern-matcher that scans protein or nucleotide sequences (DNA, RNA, and tRNA) to provide UTR motifs. It finds the motifs that identifiy 3'UTR and 5'UTR sequences on a specific sequence. Every UTR site input has been generated based on the information stated in the literature and reviewed by scientists who work experimentally on the functional characterization of related UTR items. UTRsite is a very useful database in order to find non-detected signals in given gene sequences. In this study, UTR motifs were found by the UTRscan program from "http://itbtools.ba.itb.cnr.it/utrscan" web address.

Prediction of Post-translational Modification Sites

UbPred is a database randomly predicts potential ubiquitination sites on proteins. The accuracy rate is about 72-80%. In this study, UbPred software from "www.ubpred.org" web address and BDM-PUB software from "bdmpub.biocuckoo.org" web address were employed for prediction of ubiquitination sites. 0.62-scored lysine residues were accepted as ubiquitinylated in UbPred software and balance cut-off was chosen in BDM-PUP software.

Prediction of Positions of SNPs on Different Protein Domains by Using Prosite Database

PROSITE is a protein database used for describing the functional characterization of proteins. It identifies protein domains, families and functional regions, in addition to these, their associated patterns and profiles have been shown. It provides additional information about functionally and/or structurally critical amino acids, and it is completed with ProRule which is a group of rules relying on profiles and patterns and enhances the characteristic potential of these profiles and patterns. PROSITE is mostly prefered for the description of domain properties of UniProtKB/Swiss-Prot entries. In this study, "https://prosite.expasy.org/" web page was used to predict the functional characterization of protein domains.

RESULTS

TRAF1, a member of TNF receptor (TNFR) associated factor (TRAF) protein family 1, is encoded by TRAF1 gene. Signal transduction from multiple receptors of the TNFR superfamily are madiated by TRAF proteins. On the other hand, TRAF2 has a direct interaction with TNF receptors and builds a heterodimeric complex with TRAF1. This protein is necessary for TNF-associated activation of MAPK8/JNK and NF-kappa B. It interacts with the protein complex created by TRAF1 and the inhibitor of apoptosis proteins (IAPs). It is a mediator of anti-apoptotic signals from TNF receptors. Moreover, TRAF6 employs mediating signaling from the members of the TNF receptor superfamily in addition to the Toll/IL-1 family. According to the studies, CD40, TNFSF11/RANCE, and IL-1 receptor signals are mediated with TRAF6. Also, it has been found that TRAF6 interacts with a variety of protein kinases including IRAK1/IRAK, SRC, and PKC zeta, that bridge over different signaling pathways. It functions as a signal transducer in the NF-kappaB pathway that activates the IkappaB kinase (IKK) in response to proinflammatory cytokines. Other genes that interact with the CD40 by the GeneMANIA program are shown in **Figure 1**.

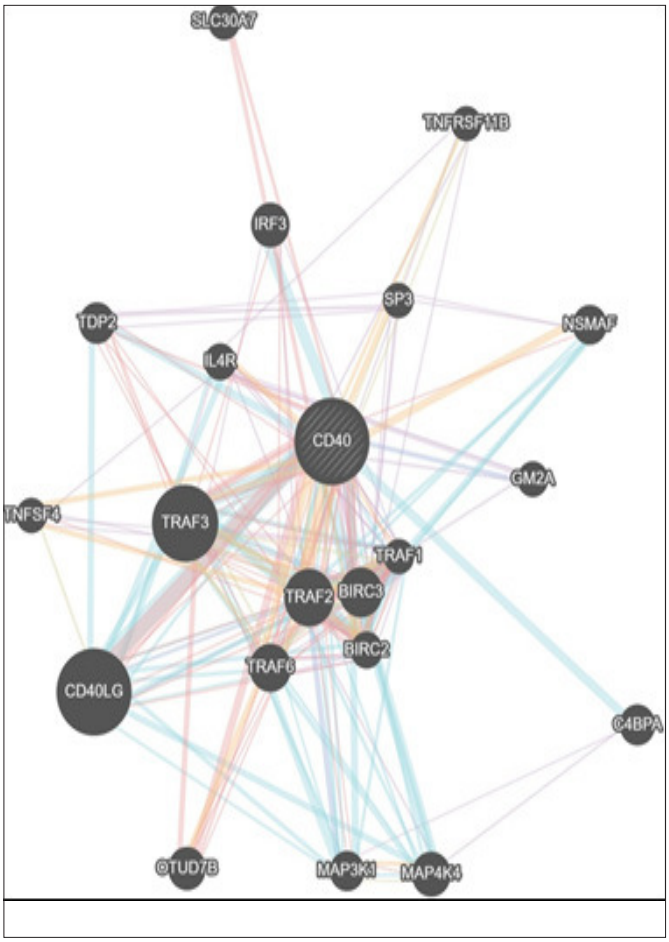

Figure 1. Genes that CD40 have physical interactions (pink) and co-expressions (blue) are shown

A total of 85 SNPs, 44 SNPs in the EA population and 41 SNPs in the AA population, were found by using Exome Variant Server (**Table 3**), and these SNPs were evaluated (**Table 1**). All SNPs were analyzed and 14 SNPs without any ClinVar data which are thought to be suspected and have not been investigated by researches yet were detected (**Table 4**). Detection of SNP Effects on miRNAs by Using mrSNP Software was shown in **Table 5**. **Table 6** has shown that associations with human diseases and evaluation of three SNPs found on CD40 gene in **Table 7**. Untranslated regions on SNPs was identified by UTRscan in **Table 8**. Potential ubiquitination sites on proteins has shown by UpBred (**Table 9**). Finally; we described the functional characterization of proteins in **Table 10**.

Table 1. Detection of polymorphisms on CD40 genes by using Exome Serve Variant Software

Variant GRCh37 Pos	rs ID	Alleles	All Allele #	All Genotype #	GVS Function	Conservation (GERP)	PolyPhen2 (Class:Score)
20:44746913	rs11569300	R>A1	A1=168/ R=12338	A1A1=5/A1R=158/RR=6090	utr-5	-0.75	unknown
20:44746942	rs11569301	C>T	T=458/ C=12548	TT=23/TC=412/CC=6068	utr-5	-3.74	unknown
20:44746946	rs373167365	G>C	C=2/ G=13004	CC=0/CG=2/GG=6501	utr-5	0.01	unknown
20:44746950	rs377285521	C>T	T=1/ C=13005	TT=0/TC=1/CC=6502	utr-5	1.33	unknown
20:44746963	rs371166508	C>G	G=3/ C=13003	GG=1/GC=1/CC=6501	utr-5	1.94	unknown
20:44746982	rs1883832	T>C	C=10444/ T=2562	CC=4255/CT=1934/TT=314	utr-5	0.76	unknown
20:44747004	rs113207193	T>G	G=3/ T=13003	GG=0/GT=3/TT=6500	missense	2.77	probably-damaging:1.0
20:44747049	rs375622419	C>T	T=1/ C=13005	TT=0/TC=1/CC=6502	intron	1.68	unknown
20:44747086	rs745307	G>A	A=892/ G=3674	AA=112/AG=668/GG=1503	intron	-0.69	unknown
20:44747104	rs11569302	C>T	T=128/ C=4438	TT=3/TC=122/CC=2158	intron	-3.37	unknown
20:44750424	rs11569315	C>T	T=111/ C=12895	TT=4/TC=103/CC=6396	intron	0.3	unknown
20:44750444	rs187683423	C>T	T=8/ C=12998	TT=1/TC=6/CC=6496	intron	2.45	unknown
20:44750480	rs147677886	G>T	T=4/ G=13002	TT=0/TG=4/GG=6499	missense	-3.65	benign:0.007
20:44750594	rs374320594	C>T	T=1/ C=4565	TT=0/TC=1/CC=2282	intron	-6.89	unknown
20:44750829	rs371933529	T>A	A=2/ T=13004	AA=0/AT=2/TT=6501	intron	-7.43	unknown
20:44750831	rs374336021	G>C	C=1/ G=13005	CC=0/CG=1/GG=6502	intron	3.79	unknown
20:44750850	rs11569317	C>G	G=815/ C=12191	GG=31/GC=753/CC=5719	intron	0.08	unknown
20:44750945	rs142258778	A>G	G=3/ A=13003	GG=0/GA=3/AA=6500	coding-synonymous	-4.43	unknown
20:44750948	rs371950759	C>A	A=1/ C=13005	AA=0/AC=1/CC=6502	missense	0.28	benign:0.352
20:44750980	rs376829285	A>G	G=1/ A=13005	GG=0/GA=1/AA=6502	missense	3.45	probably-damaging:1.0
20:44751023	rs369889788	G>A	A=2/ G=13004	AA=0/AG=2/GG=6501	intron	-7.51	unknown
20:44751040	rs61760051	A>G	G=8/ A=12998	GG=0/GA=8/AA=6495	intron	-4.32	unknown
20:44751047	rs201089032	C>G	G=20/ C=12986	GG=0/GC=20/CC=6483	intron	1.42	unknown
20:44751048	rs369286250	G>A	A=1/ G=13005	AA=0/AG=1/GG=6502	intron	-0.02	unknown
20:44751217	rs199588140	G>A	A=6/ G=13000	AA=0/AG=6/GG=6497	intron	-0.5	unknown
20:44751226	rs376448796	C>T	T=2/ C=13004	TT=0/TC=2/CC=6501	intron	1.9	unknown
20:44751229	unknown	R>A1	A1=16/ R=12502	A1A1=7/A1R=2/RR=6250	intron	0.21	unknown

Table 1. Detection of polymorphisms on CD40 genes by using Exome Serve Variant Software (cont...)

Variant GRCh37 Pos	rs ID	Alleles	All Allele #	All Genotype #	GVS Function	Conservation (GERP)	PolyPhen2 (Class:Score)
20:44751241	rs369901991	C>T	T=1/ C=13005	TT=0/TC=1/CC=6502	intron	3.89	unknown
20:44751260	rs144542285	C>T	T=1/ C=13005	TT=0/TC=1/CC=6502	missense	2.94	benign:0.376
20:44751337	rs147816161	C>T	T=1/ C=13005	TT=0/TC=1/CC=6502	coding-synonymous	-1.71	unknown
20:44751363	rs11569321	C>T	T=566/ C=12440	TT=37/TC=492/CC=5974	missense	-2.5	benign:0.012
20:44751370	rs150312341	G>T	T=1/ G=13005	TT=0/TG=1/GG=6502	coding-synonymous	-3.23	unknown
20:44751393	rs371997367	T>C	C=1/ T=13005	CC=0/CT=1/TT=6502	missense	1.73	possibly-damaging:0.651
20:44751415	rs11699100	A>G	G=283/ A=12723	GG=6/GA=271/AA=6226	intron	2.12	unknown
20:44751438	rs377499066	G>A	A=1/ G=13005	AA=0/AG=1/GG=6502	intron	1.55	unknown
20:44751719	rs372604011	G>A	A=1/ G=13005	AA=0/AG=1/GG=6502	intron	2.04	unknown
20:44751790	rs373653555	C>T	T=1/ C=13005	TT=0/TC=1/CC=6502	coding-synonymous	-6.12	unknown
20:44751796	rs376668410	C>T	T=1/ C=13005	TT=0/TC=1/CC=6502	coding-synonymous	0.19	unknown
20:44751835	rs79661585	C>T	T=1/ C=13005	TT=0/TC=1/CC=6502	coding-synonymous	-2.5	unknown
20:44755264	rs370003801	C>T	T=1/ C=13005	TT=0/TC=1/CC=6502	intron	-2.16	unknown
20:44755279	rs144600981	C>G	G=1/ C=13005	GG=0/GC=1/CC=6502	missense-near-splice	0.26	benign:0.041
20:44755279	rs144600981	C>G	G=1/ C=13005	GG=0/GC=1/CC=6502	intron	0.26	unknown
20:44755312	rs374572404	A>G	G=1/ A=13005	GG=0/GA=1/AA=6502	coding-synonymous	-3.33	unknown
20:44755312	rs374572404	A>G	G=1/ A=13005	GG=0/GA=1/AA=6502	intron	-3.33	unknown
20:44755376	rs41282788	G>C	C=188/ G=12818	CC=2/CG=184/GG=6317	intron	-2.01	unknown
20:44756742	rs193080413	G>A	A=25/ G=12981	AA=0/AG=25/GG=6478	intron	-1.05	unknown
20:44756751	rs377180915	G>A	A=1/ G=13005	AA=0/AG=1/GG=6502	intron	-5.89	unknown
20:44756823	rs7273698	C>T	T=587/ C=12419	TT=41/TC=505/CC=5957	missense	-1.19	possibly-damaging:0.894
20:44756823	rs7273698	C>T	T=587/ C=12419	TT=41/TC=505/CC=5957	coding-synonymous	-1.19	unknown
20:44756827	rs143037975	A>G	G=3/ A=13003	GG=0/GA=3/AA=6500	missense	-5.67	benign:0.0
20:44756890	rs11569337	G>T	T=133/ G=12873	TT=2/TG=129/GG=6372	intron	0.35	unknown
20:44756891	rs3765456	G>A	A=1482/ G=11524	AA=82/AG=1318/GG=5103	intron	2.66	unknown
20:44756903	rs11569338	A>T	T=17/ A=12989	TT=0/TA=17/AA=6486	intron	-6.71	unknown
20:44756908	rs4813000	A>C	C=83/ A=12923	CC=0/CA=83/AA=6420	intron	1.46	unknown
20:44756910	rs374618623	C>T	T=1/ C=13005	TT=0/TC=1/CC=6502	intron	-0.86	unknown
20:44756977	rs368619894	G>A	A=1/ G=13005	AA=0/AG=1/GG=6502	missense	-6.62	benign:0.004
20:44756981	rs371799172	A>G	G=1/ A=13005	GG=0/GA=1/AA=6502	missense	0.15	benign:0.015
20:44756981	rs371799172	A>G	G=1/ A=13005	GG=0/GA=1/AA=6502	coding-synonymous	0.15	unknown
20:44757016	rs368614125	C>T	T=1/ C=13005	TT=0/TC=1/CC=6502	intron	-7.28	unknown

Table 1. Detection of polymorphisms on CD40 genes by using Exome Serve Variant Software (cont...)

Variant GRCh37 Pos	rs ID	Alleles	All Allele #	All Genotype #	GVS Function	Conservation (GERP)	PolyPhen2 (Class:Score)
20:44757018	rs73310754	C>T	T=66/ C=12940	TT=0/TC=66/CC=6437	intron	-100.0	unknown
20:44757028	rs368055606	T>C	C=1/ T=13005	CC=0/CT=1/TT=6502	intron	-0.87	unknown
20:44757475	rs11697349	C>T	T=268/ C=12736	TT=5/TC=258/CC=6239	intron	-4.17	unknown
20:44757524	rs11086998	C>G	G=86/ C=12920	GG=0/GC=86/CC=6417	missense	-4.88	benign:0.003
20:44757524	rs11086998	C>G	G=86/ C=12920	GG=0/GC=86/CC=6417	utr-3	-4.88	unknown
20:44757526	rs148342289	C>T	T=6/ C=13000	TT=0/TC=6/CC=6497	coding-synonymous	-1.96	unknown
20:44757526	rs148342289	C>T	T=6/ C=13000	TT=0/TC=6/CC=6497	utr-3	-1.96	unknown
20:44757562	rs144466131	C>T	T=18/ C=12988	TT=0/TC=18/CC=6485	coding-synonymous	-6.45	unknown
20:44757562	rs144466131	C>T	T=18/ C=12988	TT=0/TC=18/CC=6485	utr-3	-6.45	unknown
20:44757595	rs376780996	A>G	G=1/ A=13005	GG=0/GA=1/AA=6502	coding-synonymous	-9.6	unknown
20:44757595	rs376780996	A>G	G=1/ A=13005	GG=0/GA=1/AA=6502	utr-3	-9.6	unknown
20:44757622	rs369693842	A>G	G=3/ A=13003	GG=0/GA=3/AA=6500	coding-synonymous	2.65	unknown
20:44757622	rs369693842	A>G	G=3/ A=13003	GG=0/GA=3/AA=6500	utr-3	2.65	unknown
20:44757625	rs150890139	G>A	A=1/ G=13005	AA=0/AG=1/GG=6502	coding-synonymous	-9.23	unknown
20:44757625	rs150890139	G>A	A=1/ G=13005	AA=0/AG=1/GG=6502	utr-3	-9.23	unknown
20:44757654	rs139300926	G>A	A=1/ G=13005	AA=0/AG=1/GG=6502	missense	4.61	probably-damaging:1.0
20:44757654	rs139300926	G>A	A=1/ G=13005	AA=0/AG=1/GG=6502	utr-3	4.61	unknown

Table 2. Estimating the positions of nsSNPs in different protein domains using a prosite database

Sequence	Score	Predicted Features of Disulfide	Condition
25-59 ACREKQYLIN-SQC-CSLCQPGQKLVSDCTEFTETEC	9.423	26-37 38-51 41-59	C-x*-C C-x*-C C-x*-C
61-103 PCGESEFLDTWNRETHCHQHKYCDPNLg 1RVQQKGTSETDTIC	9.488	62-77 83-103	C-x*-C C-x*-C
104-143 TCEEGWHCTS----EACESCV1hrSSPGFGVKQIATGCSDTIC	9.231	105-116 125-143	C-x*-C C-x*-C
145-186 PCPVGFFSNVSSAFEKCHPWTSCERKD1 VVQQAGTNKTDVVC	8.995	146-161 167-186	C-x*-C C-x*-C

Table 3. The number of variations of CD40 gene on European-American and African-American populations.

Number of Variation	Population
44	European-American
41	African-American

Table 4. Evaluation of suspected SNPs found by using Exome Serve Variant Software

rs ID	Report
rs113207193	Not reported in ClinVAR
rs147677886	Not reported in ClinVAR
rs371950759	Not reported in ClinVAR
rs379829285	Not reported in ClinVAR
rs144542285	Not reported in ClinVAR
rs11569321	Reported in ClinVAR
rs371997367	Not reported in ClinVAR
rs144600981	Not reported in ClinVAR
rs7273698	Reported in ClinVAR
rs143037975	Not reported in ClinVAR
rs368619894	Not reported in ClinVAR
rs371799172	Not reported in ClinVAR
rs11086998	Reported in ClinVAR
rs139300926	Not reported in ClinVAR

Table 5. SNPs and INDELs in mIRNA target sites

Location	dbSNP ID	Wobble base pair	Ancestral Allele	Allele	mİR ID	mİRSite
44757565	rs147672904	N	C	T	hsa-mIR-5694	tcccgATGATCTt
44757599	rs199980487	N	C	C	hsa-mIR-1976	ccagtGCAGGAGA
					hsa-mIR-4693-3p	ccAGTGCAGgaga
				T	hsa-mIR-155-3p	ccagTGTAGGAga
					hsa-mIR-6757-3p	CCAGTTGTAggaga
44757625	rs150890139	Y	G	A	hsa-mIR-134-5p	cccaacCAGTCACc
					hsa-mIR-3118	CCAACAgtcacc
					hsa-mIR-92a-1-5p	CCCAGTGTAgggaga
44767654	rs139300926	Y	G	G	hsa-mIR-4431	ccaacCAGTCACc
					hsa-mIR-4708-5p	agAGTCACAtctc
				A	hsa-mIR-134-5p	agAGTCACAtctc
					hsa-mIR-3118	agAGTCACAtctc
					hsa-mIR-3164	agAGTCACAtctc
					hsa-mIR-4501	agaGTCACATctc
					hsa-mIR-4522	AGAGTCAcatctc
					hsa-mIR-576-3p	agagtCACATCTc
					hsa-mIR-6820-3p	agAGTCACAtctc

Table 6. Associations with human diseases and traits

Disease/Trait	Pubmed ID	Marker ID	Study	P-Value
Multiple Sclerosis	21833088	rs2425752	Genetic risk and primary role for cell-mediated immune mechanisms in multiple sclerosis.	5E-10
Multiple Sclerosis	22190364	rs6074022	Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci.	5E-6
Multiple Sclerosis	19525955	rs6074022	Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20.	1E-7
Inflammatory Bowel Diseases	23128233	rs1569723	Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.	1E-13
Kawasaki disease	22446961	rs1569723	Two new susceptibility loci for Kawasaki disease identified through genome-wide association analysis.	6E-9
Rheumatoid Arthritis	20453842	rs4810485	Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci.	3E-9
Rheumatoid Arthritis	18794853	rs4810485	Common variants at CD40 and other loci confer risk of rheumatoid arthritis.	8E-9
Rheumatoid Arthritis	22446962	rs4810485	A genome wide association study identifies three new risk loci for Kawasaki disease.	6E-8

Table 7. Evaluation of three SNPs found on CD40 gene

Pathway ID	Class	Description
Hsa05416	Cardiovascular diseases	Viral myocarditis
Hsa05310	Immune diseases	Asthma
Hsa05320	Immune diseases	Autoimmune thyroid disease
Hsa05322	Immune diseases	Systemic lupus erythematosus
Hsa05330	Immune diseases	Allograft rejection
Hsa05340	Immune diseases	Primary immunodeficiency
Hsa04620	Immune diseases	Toll-like receptor signaling pathway
Hsa04672	Immune diseases	Intestinal immune network for IgA production
Hsa04060	Signaling molecules and interaction	Cytokine-cytokine receptor interaction
Hsa04514	Signaling molecules and interaction	Cell adhesion molecules (CAMs)

Table 8. Detecting SNPs on untranslated regions by the UTRscan program

Number	Name	Number	Name
1	Histone 3'UTR stem-loop structure (HSL3)	17	UNR binding site (UNR-BS)
2	Iron Responsive Element (IRE)	18	Ribosomal S12 mitochondrial protein 5'UTR translation control element (RPMS12_TCE)
3	Seleno cysteine Insertion Sequence - type 1 (SCIS1)	19	Bruno 3'UTR responsive element (BRE)
4	Seleno cysteine Insertion Sequence - type 2 (SCIS2)	20	Alcohol dehydrogenase 3'UTR down-regulation control element (ADH_DRE)
5	Amyloid precursor protein mRNA stability control element (APP_SCE)	21	Barley yellow dwarf virus translation control element (BYDV_TE)
6	Cytoplasmic polyadenylation element (CPE)	22	Proneural Box (PB)
7	Translational regulation element (TGE)	23	Brd-Box (Brd)
8	Nanos translation control element (NANOS_TCE)	24	K-Box (K-BRD)
9	15-Lipoxygenase differentiation Control Element (15-LOX-DICE)	25	Gy-Box (GY)
10	AU-rich class-2 Element (ARE2)	26	Androgen receptor CU-rich element (AR_CURE)
11	Terminal Oligopyrimidine Tract (TOP)	27	Elastin G3A 3'UTR stability motif (G3A)
12	Glusosetransporter type-1 3'UTR cis-acting element (GLUT1)	28	Insulin 3'UTR stability element (INS_SCE)
13	Tumor necrosis factor alpha 3'UTR cis-acting element (TNF)	30	Beta-actin 3'UTR zipcode (ACTIN_ZIP3)
14	Vimentin 3'UTR cis-acting element (VIM3)	31	Gap-43 Stabilization Element (GAP-43)
15	Internal Ribosome Entry Site (IRES)	32	Gamma interferon activated inhibitor of translation (GAIT element)
16	SXL binding site (SXL_BS)		

Table 9. Estimating protein ubiquitination regions

Residue	Score	Ubiquitinated
29	0.59	No
46	0.43	No
81	0.60	Yes, medium confidence
94	0.78	No
132	0.50	No
160	0.51	No
170	0.51	No
181	0.68	Yes, low confidence
216	0.45	No
217	0.41	No
220	0.33	No
221	0.24	No
225	0.23	No
230	0.80	Yes, high confidence
267	0.81	Yes, high confidence

(Default low confidence according to the program; score range 0.62 <s <0.69, sensitivity 0.464, specificity 0.903, medium confidence; score range 0.69 <s <0.84, sensitivity 0.346, specificity 0.950/high confidence; score range 0.84 <s <1.00, sensitivity 0.197, specificity 0.989).

Table 10. Estimation of ubiquitination regions.

Peptide	Position	Score	Threshold
PPTACREKQYLINSQ	29	1.62	0.3
SLCQPGQKLVSDCTE	46	2.46	0.3
LGLRVQQKGTSETDT	94	0.99	0.3
NVSSAFEKCHPWTSC	160	0.84	0.3
PWTSCETKDLVVQQA	170	0.64	0.3
LLVLVFIKKVAKKPT	216	1.37	0.3
VFIKKVAKKPTNKAP	220	1.74	0.3
FIKKVAKKPTNKAPH	221	2.33	0.3
VAKKPTNKAPHPKQE	225	1.67	0.3
PVTQEDGKESRISVQ	267	1.10	0.3

DISCUSSION

The allele frequency of a polymorphic gene can vary geographically and ethnically (19). Polymorphic changes related to SNPs found on genes can also change the response and treatment process of diseases. Therefore, detecting population-specific polymorphic changes will help the physician in diagnosis and provide an advantage for treatment. CD40 is an important protein that is a member of the TNF family. Previous studies have revealed that TRAF1, TRAF2, TRAF6 proteins interact with the CD40 receptor and act as mediators of signal transduction. In addition to these genes, according to the data we obtained from our in silico results, it has been demonstrated via the GeneMANIA program that CD40 also has interaction with some other genes such as SP3, TDP2, IL4R, GM2A, NSMAF, BIRC3, BIRC2, OTUD78, MAP3K1, SLC30A7, MAP4K4. In future studies, these data can provide an estimate of the contribution of CD40 to other diseases. 14 missense variants from 85 SNPs of CD40 have been found via the Exome Variant Server program. Only 3 of these SNPs were identified in ClinVar. In the literature, these SNPs have not been studied experimentally before. There is no information in the literature about 11 other variants. Therefore, the relationship of SNPs with diseases can contribute to population-specific studies with new experimental studies. In the population-specific study, the population-specific gene mapping can be generated by designing primers and probes for each SNPs that are found and scanning mutations with RT-PCR. In our study, according to the data from PredictSNP program,

rs139300926 SNP is likely to have a mutation in the amino acid 270. This data has also been confirmed by PinSNPs. Our study especially attracted attention that the variants rs147677886, rs11569321, rs7273698, rs11086998 and rs139300926 are suspicious. The data in **Table 8** were analyzed to show the effect of binding of 3'UTR SNP on miRNA by mrSNP. Suspected 14 missense variants were run at dbSNP and according to the results, rs139300926, rs147672904, rs199980487, rs150890139 variants were defined. Among these SNPs, rs150890139 has two different definitions as unknown in Exome Variant Server, rs199980487 is defined as "unknown", and rs147672904 and rs139300926 were defined as "suspicious SNP" in the same server. There are no experimental studies in the literature with these SNPs. Our bioinformatics study indicates the fact that these SNPs have a promising future experimentally.

Single nucleotide polymorphisms (SNPs) have been detected in 3'UTRs that break down normal miRNA binding or create new binding sites, some of which have been correlated with disease pathogenesis. That increases the significance of detecting miRNA targets and predicting the potential effects of SNPs on binding sites. In this study, we have provided very detailed information about 4 SNPs and their association with diseases. Related SNPs have been associated with Multiple Sclerosis (MS), Rheumatoid Arthritis (RA), and Kawasaki Disease.

The effect of untranslated regions on transcriptional motifs can be estimated via the UTRscan program (20). Untranslated regions play an important role in the regulation of gene expression after transcription, in the stability and efficiency of translation. In our study, 32 untranslated regions were found by UTRscan prediction. UbPred is a web tool that determines the possibility of whether a lysine residue in a protein is ubiquitin (21). The results signaled at the residues of 81, 181, 230, and 267 (**Table 9**). According to the data obtained from the Prosite database (**Table 2**), rs772829518, rs779766201 mutations were found in residues 116 and 167, respectively. These mutations were checked whether they belong to CD40 via PinSnps ve U.S. National Library of Medicine. The position of rs772829518 was detected on chr20:46122699 (GRCh38.p12) and it is a missense variant. Moreover, the position of rs779766201b was detected on chr20:46126642 (GRCh38.p12) and it is also a missense variant.

CONCLUSION

Finally, in future studies, the three-dimensional structure of CD40 could be generated and the locations of these mutations on the protein can be identified.

ETHICAL CONSIDERATIONS

Ethics Committee Approval: In this study, in silico methods have been used and do not require any ethical committee approval.

Informed Consent: In this study, in silico methods have been used, do not have any biological material, therefore do not require informed consent.

Referee Evaluation Processs: Externally peer-reviewed.

Conflict of Interest Statement: The authors have no conflicts of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

Author Contributions: All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version

REFERENCES

1. Kooten C, Banchereau J. Functions of CD40 on B cells, dendritic cells and other cells. Curr Opin Immunol 1997; 9: 3: 330-7
2. Kawabe T, Matsushima M, Hashimoto N, et al. CD40/CD40 ligand interactions in immune responses and pulmonary immunity. Nagoya J MedSci 2011; 73: 3: 69-8.
3. Elgueta R, Micah J. Benson, et al. Molecular mechanism and function of CD40/CD40L engagement in the immune system. Immunol Rev 2009; 229: 1.
4. Davies C, Mak W, Young S, et al. TRAF6 Is Required for TRAF2-Dependent CD40 signal transduction in nonhemopoietic cells. Mol Cell Biol 2005; 25: 22: 9806-19.
5. Ghezala A, Frieling J, Paris D, et al. CD40 promotion of amyloid beta production occurs via the NF-kappaB pathway. Eur J Neurosci 2007;25: 6: 1685-95.
6. Charles AJ, Travers P, Walport M, et al. Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science; 2001. ISBN-10: 0-8153-3642-X.
7. Kawabe T, Naka T, Yoshida K, et al. The immune responses in CD40-deficient mice:impaired immunoglobulin classs witching and germinal center formation. Immunity 1994; 1: 3: 167-8.
8. Chatzigeorgiou A, Lyberi M, Chatzilymperis G, et al. CD40/CD40L signaling and its implication in health and disease". Biofactors 2009; 35: 6: 474-3.
9. Carlring J, Altaher HM, Clark S, et al. CD154-CD40 interactions in the control of murine B cell hematopoiesis. J Leukoc Biol 2011; 89: 5: 697-06.
10. Zhang SQ, Han JW, Sun LD, et al. A single-nucleotide polymorphism of the TNFSF4 gene is associated with systemic lupus erythematosus in Chinese Han Population. Rheumatol Int 2011; 31: 227-31.
11. Zhu Y, Wang J, Feng X. CTLA-4 SNPs (CT60A/G, -1722T/C, -1661G/A, and -318C/T) and systemic lupus erythematosus: a meta-analysis. Crit Rev Eukaryot Gene Expr 2014; 24: 89-100.
12. Orozco G, Eyre S, Hinks A, et al. Association of CD40 with rheumatoid arthritis confirmed in a large UK case-control study. Ann Rheum Dis 2010; 69: 813-16.
13. Raychaudhuri S, Remmers EF, Lee AT, et al. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat Genet 2008; 40: 1216-23.

14. Burdon KP, Langefeld CD, Beck SR, et al. Variants of the CD40 gene but not of the cd40l gene are associated with coronary artery calcification in the diabetes heart study (DHS). Am Heart J 2006; 151: 706-11.

15. Chen F, Hou S, Jiang Z, et al. CD40 polymorphisms in han chinese patients with fuch uveitis syndrome. Mol Vis 2011; 17: 2469-72

16. Yan JC, Ma GS, Wu ZG et al. ıncreased levels of CD40-CD40 ligand system in patients with essential hypertension. Clin Chim Acta 2005; 355: 191-6.

17. Li M, Sun H, Liu S, et al. CD40 C/T-1 polymorphism plays different roles in graves' disease and hashimoto's thyroiditis: a meta-analysis. Endocr J 2012; 59: 1041-50.

18. Zhang B, Wu T, Song C, et al. Association of CD40–1C/T polymorphism with cerebral ınfarction susceptibility and its effect on CD40L in chinese population. Int Immuno pharmacol 2013; 16: 461-5.

19. Büküm E, Düzgün A, Karkucak M, et al. Mide kanserli hastalarda siklin D1 (G870A) gen polimorfizminin araştırılması. Konuralp Tıp Derg 2013; 5: 1: 18-2.

20. Li B, Krishnan VG, Mort ME, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. 2009; 25: 21: 2744-0.

21. Radivojac, P, Vacic V, Haynes C, et al. Identification analysis and prediction of protein ubiquitination sites. Proteins. 2010; 78: 2: 365-80.